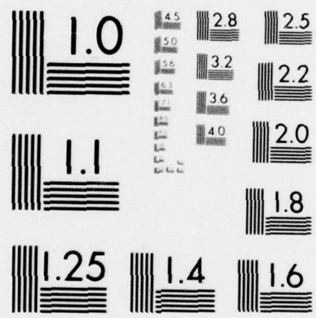
AD-A072 590 WISCONSIN UNIV-MADISON DEPT OF STATISTICS F/G 12/1
STATISTICAL MODELS FOR TIME SERIES AND LIFE TESTING WITH APPLIC--ETC(U)
1978 G E BOX, G K BHATTACHARYYA AFOSR-77-3272
UNCLASSIFIED AFOSR-TR-78-1453 NL

| OF |

AD
A072590

END
DATE
FILMED
9-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- 78-1453

LEVEL III

A041766

AD A072590

FINAL REPORT

Air Force Office of Scientific Research

Grant AF-AFOSR 77-3272

1 April 1977 - 30 June 1978

G. E. P. Box
Principal Investigator

G. K. Bhattacharyya
Richard A. Johnson
Grace Wahba

Associate Investigators

Approved for public release;
distribution unlimited.

19 Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 18 AFOSR-TR-78-1453
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): 6 STATISTICAL MODELS FOR TIME SERIES and LIFE TESTING WITH APPLICATIONS IN ENGINEERING SYSTEMS.
5. TYPE OF REPORT & PERIOD COVERED: 9 Final rept. 1 Apr 77 - 30 Jun 78
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): 10 George E.P. Box, G.K. Bhattacharyya, Richard A. Johnson [redacted] Grace Wahba
8. CONTRACT OR GRANT NUMBER(s): 15 AFOSR-77-3272
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Wisconsin-Madison, Department of Statistics, Madison, Wisconsin 53706 ok
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 16 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332
12. REPORT DATE: 11 1978 17 A5
13. NUMBER OF PAGES: 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 14p.
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
18. SUPPLEMENTARY NOTES
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

→ Over the period of this grant 5 papers have been accepted for publication, 1 paper has been published, 5 technical reports have been written, and 4 theses have been completed. Over the period 1 April 1972 to June, 1978 35 papers have been published, 9 papers have been accepted for publication, 41 technical reports have been written, one book has been written and another revised, all of these have received support from the present and preceeding Air Force grants. Many of the publications have reflected our continuing interest in the topics of stochastic systems (especially with

DD FORM 1 JAN 73 1473

400 243 DW

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. Abstract continued.

discrete control) in model building, smoothing and curve estimation, methods of approximation with noisy data, inferences with censored data, and life testing and reliability with applications to systems.

Accession For	
NTIS GRA&I	☒
DDC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist.	Avail and/or special
A	

UNCLASSIFIED

STATISTICAL MODELS FOR TIME SERIES AND LIFE TESTING WITH APPLICATIONS IN ENGINEERING SYSTEMS

OVERVIEW OF RESEARCH

Over the period of this grant 5 papers have been accepted for publication, 1 paper has been published, 5 technical reports have been written, and 4 theses have been completed. Over the period 1 April 1972 to June, 1978 35 papers have been published, 9 papers have been accepted for publication, 41 technical reports have been written, one book has been written and another revised, all of these have received support from the present and preceeding Air Force grants. Many of the publications have reflected our continuing interest in the topics of stochastic systems (especially with discrete control) in model building, smoothing and curve estimation, methods of approximation with noisy data, inferences with censored data, and life testing and reliability with applications to systems.

Part 1 - Time Series Models

Research in this contract period has led to a deeper understanding of stochastic models containing deterministic components, and of the relationship between methods of forecasting using ARMA models and generalized exponential smoothing. Also the question of optimal sampling periods for efficient sample data feedback control has received attention.

Deterministic or Stochastic Models?

Models of the form

$$y_t = f(t) + \varepsilon_t$$

in which $f(t)$ represents a deterministic function of time and ε_t a random error have often been misused to represent series whose development could much more plausibly be represented by a stochastic model such as an autoregressive-moving average ARMA process.

There will, however, be situations in which a deterministic component actually exists. Our research shows that appropriate analysis of a fitted ARMA process can point to the necessity for a deterministic component.

A paper describing these findings has recently been accepted for publication in "Applied Statistics."[1]

When are exponential smoothing methods optimal?

Generalized exponential smoothing forecasting procedures are used extensively in many areas of economics, business and engineering. Our research shows that:

i) These forecasting procedures are optimal in terms of achieving minimum mean square error forecasts only if the underlying stochastic process is included in a limited subclass of ARIMA (p,d,q) processes. Hence, it is shown what assumptions are made when using these procedures.

ii) The implication of point (i) is that the users of these procedures tacitly assume that the stochastic processes which occur in the real world are from a very restricted subclass of stochastic processes. No reason can be found why these particular models should occur more frequently than others.

iii) It is further shown that even if a stochastic process which would lead to such a procedure did occur the actual methods used for making the forecasts are clumsy. Much simpler methods are available.

A paper describing these findings was recently published in "Metrika."[2]

Sampling Interval and Feedback Control

One question which often arises in sample data feedback systems is how frequently should one sample. In an optimal control scheme, we suppose that a control arrangement will be employed in which the sampling interval is h units long where h is an unknown integer. The equations describing both the dynamics and the noise process now depend on this interval and a reasonable cost function may be formulated as a function of h. The optimum time interval at which surveillance should be conducted, i.e. observations taken and control adjustment action taken, is obtained by minimizing the above cost function.

A paper describing these findings has been accepted for publication in "Technometrics."[3]

Spline Smoothing for the Non-Parametric Recovery of Curves

Consider the model

$$Y(t_i) = g(t_i) + \varepsilon_i, \qquad i = 1,2,\ldots,n \qquad t_i \in T$$

where $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)' \sim N(0,\sigma^2 I_{n\times n})$ and $g(\cdot)$ is some "smooth" function defined on some index set T. When T is an interval of the real line, cubic polynomial smoothing splines are well known to provide an esthetically satisfying method for estimating $g(\cdot)$, from a realization $\underset{\sim}{y} = (y_1,\ldots,y_n)'$ of $\underset{\sim}{Y} = (Y(t_1),\ldots,Y(t_n))'$. Splines are an appealing alternative to fitting a specified set of m regression functions, for example polynomials of degree less than m, when one is uncertain that the true curve $g(\cdot)$ is actually in the span of the specified regression functions. We show that polynomial spline (respectively generalized spline) smoothing is equivalent to Bayesian estimation with a prior on g which is "diffuse" on the coefficients of the polynomials of degree $\leq m$ (respectively specified set of m regression functions), and "proper" over an appropriate set of random variables not including the coefficients of the regression functions. Since Gauss Markov estimation is equivalent to Bayesian estimation with a prior diffuse over the coefficients of the regression functions, this result leads to the conclusion that spline smoothing is a (the?) natural extension of Gauss-Markov regression with m specified regression functions. It is shown that spline smoothing is an appropriate solution to the problem arising when one wants to fit a given set of regression functions to the data but one also wants to "hedge" against model errors, that is, against the possibility that the true model g is not exactly in the span of the given set of regression functions. We show that the spline smoothing approach leads to a natural measure of the deviation of the true g from the span of the regression functions, and furthermore, a good value of this measure can be estimated from the data. The estimated value of the measure is then used to control the deviation of the estimated g.

A paper describing these findings has recently been accepted for publication in the Journal of the Royal Statistical Society, Ser. B, [6].

Goodness of Fit Tests

Some basic work on the theory of k-spacings in goodness-of-fit tests has been completed. The goodness-of-fit problem is the problem of testing whether a set of independent observations $X_1, X_2, \ldots, X_n$ with unknown cumulative distribution function F actually came from a population with a specified distribution F_0. If (loosely) $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ are the ordered observations, then $S_1 = X^{(2)} - X^{(1)}$, $S_2 = X^{(3)} - X^{(2)}, \ldots, S_{n-1} = X^{(n)} - X^{(n-1)}$ are the spacings. The k-spacings are defined for $n+1 = Nk$, by $D_i = X^{(ki)} - X^{(k(i+1))}$. Some classical goodness-of-fit tests are based on statistics of the form

$$T_n = \sum_i t_n(S_i) \text{ ,}$$

and tests based on k-spacings are of the form

$$\tilde{T}_N = \sum_i \tilde{t}_N(D_i) ,$$

where t_n and $\tilde{t}_N$ are, typically, convex functions. A general theory has been developed showing that tests based on k spacings, for suitable k, have better properties than tests based on ordinary spacings. The asymptotic distribution theory of $\tilde{T}_N$ under alternatives near to F_0 has been obtained. [4]. Some substantial contributions have been made to the important problem of the behavior of k-spacings goodness-of-fit tests when some parameters in F_0 must be estimated. [5]

This work has been submitted for publication.

<u>Part 1</u>

[1] Abraham, Bovas and Box, G. E. P. (1978) Deterministic and Forecast Adaptive Time Dependent Models. (Accepted for publication in <u>Applied Statistics</u>.)

[2] Ledolter, J. and Box, G. E. P. (1978) Conditions for the Optimality of Exponential Smoothing Forecast Procedures. <u>Metrika</u>, 25, pp 77-93.

[3] Abraham, Bovas and Box, G. E. P. (1978) Sampling Interval and Feedback Control. (Accepted for publication in <u>Technometrics</u>.)

[4] del Pino, G. E., (1977) K-SPACINGS AND GOODNESS OF FIT TESTS: Pt. 3. Asymptotic distribution of spacings under a sequence of local alternatives. TR485.

[5] del Pino, G. E., (1977) K-SPACINGS AND GOODNESS OF FIT TESTS: Pt. 4. Gaussian measures and their applications to the asymptotic distribution of empirical processes. TR486.

[6] Wahba, Grace, (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression. TR508. To appear <u>J. Royal Statistical Society</u>.

Part 2 - Life Testing and Reliability.

The stress-strength model for the reliability of a single component assumes that a component, having random strength Y, is placed under a random stress X by its operating environment. In [1] we first extend this model to an s out of k system having identical components. It is assumed that the distribution of Y is unknown and the distribution of X can be either known or unknown. Under the assumption that the unknown c.d.f.'s are continuous, we obtain the UMVU estimator of reliability and another estimator based on empirical distributions. Then, employing some notions of weak convergence, we derive their limiting distributions and establish asymptotic equivalence.

According to the weakest link theory for the behavior of materials or systems, life length (or strength) will have a distribution that is closely approximated by one of the three types of extreme value distributions. By considering the class of transformations $\left(1 + \frac{(x-a)/b}{\lambda}\right)^{\lambda}$, $-\infty < \lambda < 0$ and $0 < \lambda < \infty$, we find in [2] that the problem of finding the best fitting extreme value distribution is equivalent to finding a, b, and λ that transform the observations to a standard negative exponential. The power λ selected by our procedure then identifies the appropriate extreme value distribution. Our approach of finding the posterior distribution of λ also ascertains the relative fit of the best fitting models selected from each of the three different types. The large sample behavior of the technique is studied and many examples are given.

Asymptotic normality and efficiency of the modified least squares estimators (MLSE) are studied in [3,4] in the context of some accelerated life test models. A general parametric family of life distributions, involving the scale and shape parameters, is considered where the logarithm of the scale parameter is assumed to be linearly related to the stress variables. Many of the widely used engineering models (Arrhenius, Eyring, inverse power law, generalized Eyring, etc.) are special cases of this formulation. Aside from a rigorous treatment of the limiting normality of the MLSE and the maximum likelihood estimators, the asymptotic efficiencies of the MLSE are derived both under complete samples and type II censored samples. Numerical evaluations are performed for the exponential, Weibull and gamma distributions.

A semi-Markov process is formulated in [5] for a purchasing model where the inverse Gaussian distribution is used for the interpurchase times. To account for population heterogeneity, a natural conjugate prior for the model parameters is developed and the compounding distribution is fitted to panel data. A number of important summary measures are derived. These include the market share of a specified brand as a function of time, the long run behavior of the interval transition probabilities and the market share, and the probability distribution of the number of purchases in a given time span.

Semi-Markov process models for wearout have proven valuable in the determination of maintainance and replacement schedules for many complex hardware systems. In his thesis [6], M. Akritas studies inference procedures for continuous time stochastic processes. Optimal statistical procedures are derived for situations where the series is observed from $(0,t)$ and $t \to \infty$. The results obtained are quite general. Markov processes and Semi-Markov processes are treated in detail. An outgrowth of this thesis work is the related work [7] giving general conditions for contiguity.

In epidemiological investigations it is often necessary to estimate the infection rate in the population of disease transmitting agents. Because the number of specimens in the sample is frequently high, it is practically impossible to assay each specimen individually. Instead, the specimens are randomly divided into a number of "pools" and each pool is tested as a unit.

In [8] we derive the confidence interval for the infection rate when the data come from pool testing. We also derive point estimates for the situation where the finite sample itself is considered to be the population under study.

Part 2

[1] Bhattacharyya, G. K. and Johnson, R. A. (1978) Estimation of a System Reliability by Nonparametric Techniques . (To appear Bull. Math. Soc. Greece - memorial volume).

[2] Hernandez, F. (1978) The Large Sample Behavior of Transformations to Normal or Exponential Distribution. Ph.D. thesis, University of Wisconsin (R. Johnson-major professor)

[3] Soejoeti, Z. (1978) On the Performance of the Modified Least Squares Estimators in Some Accelerated Life Testing Models Using Complete and Type II Censored Samples. Ph.D. thesis, University of Wisconsin (G. K. Bhattacharyya-major professor)

[4] Bhattacharyya, G. K. and Soejoeti, Z. (1978) Asymptotic Normality and Efficiency of Modified Least Squares Estimators in some Accelerated Life Test Models, Technical Report No. 507, University of Wisconsin.

[5] Banerjee, A. K. and Bhattacharyya, G. K. (1978) A Study of Consumer Buying Behavior through Bayesian Analysis of a Semi-Markov Process, Technical Report No. 512, University of Wisconsin.

[6] Akritas, M. (1978) Contiguity of Probability Measures Associated with Continuous Time Stochastic Processes. Ph.D. thesis, University of Wisconsin (R. Johnson-major professor)

[7] Akritas, M. Puri, M. and Roussas, G. (1978) Sample Size, Parameter Rates and Contiguity—The i.i.d. case, submitted.

[8] Bhattacharyya, G. K., Karandinos, M. and DeFoliart, G. (1978) Point Estimates and Confidence Intervals for Infection Rates Using Pooled Organisms in Epidemiological Studies. (To appear in American Journal of Epidemiology).

Personnel Supported

Principal Investigators

G. K. Bhattacharyya	Sem. I, 77-78	44.44%
George Box	June 1977	50%
	Jul. & 1/2 Aug. 1977	100%
Richard Johnson	Nov. & Dec. 1977	100%
Grace Wahba	June & Jul. 1977	100%

Research Assistants

Gina Chen	9/1/77 - 3/31/78	50%
Fabian Hernandez	9/1/77 - 1/12 noon/78	50%
	1/12 noon/78 - 3/31/78	33.33%
Michael Akritas	1/12/noon/78 - 3/31/78	16.67%
Spencer Graves	1/1/78 - 3/31/78	50%
Gregory Hudak	1/1/78 - 3/31/78	50%

Secretarial

Susan Broughton,	Technical Typist -	11/6/77 - 6/17/78	100%
Luana Cesario	Student Hourly	10/16/77- 4/1/78	222 hours